Analysis of the milliQan demonstrator data at the LHC

M. Citron

September 5, 2019

1 Introduction

The collection, calibration and analysis of the milliQan demonstrator data is described below. The major remaining tasks are:

- Rerun triple prediction from cosmic data excluding runs without magnet (1379-1425)
- Make time dependent dark rate estimate (ideally using physics runs)
- Add cosmic shower section
- Add after pulse section
- Add radiation section
- Add interpretation section
- Refine signal selection based on simulation for background and signal
- N_{PE} dependent signal selection to target higher and lower charges

2 The milliQan demonstrator

The milliQan demonstrator is positioned in the PX56 drainage gallery in the site envisioned for the full milliQan detector. PX56 has a diameter of ~ 2.7 m. With the aid of a 3D model constructed from a laser scan of the

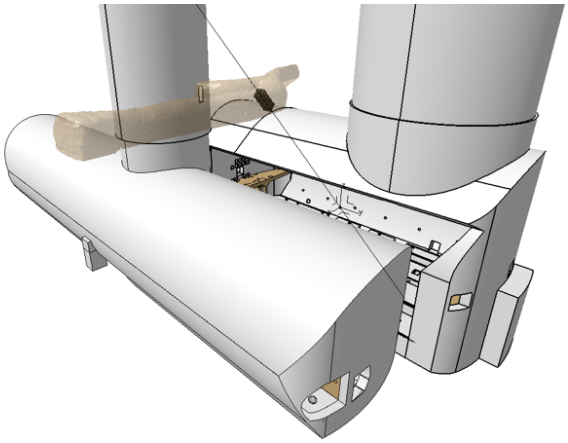


Figure 1: 3D model showing optimal position of milliQan within the PX56 Drainage and Observation gallery located above CMS UXC.

tunnel, an optimized location has been determined to place milliQan that intercepts maximal mCP flux given the constraint that, at the chosen angle to the IP, it must be able accommodate the length of the detector. The chosen location is almost directly above CMS with a distance to the IP of 33 m (17 m of which is through rock), and at an azimuthal angle of 43.1 degrees from the horizontal plane. The 3D model showing the location of the demonstrator is shown in Fig. 1.

The milliQan demonstrator is shown in Figure 2. It consists of eighteen $5 \times 5 \times 80$ cm scintillator bars arranged in three layers of 2×3 scintillator+PMT units. In addition to the bars, "slabs" of $2.5 \times 20 \times 30$ cm scintillator and thin panels of $1 \times 18 \times 100$ cm scintillator are inserted to tag charged particles such as muons from the CMS IP, study backgrounds from radiation, and to simulate the active veto of the full detector. There are also several hodoscope packs composed of small arrays of 0.75×18 inch rectangular pieces of plastic scintillator readout via SiPMs attached at one end. These provide finer grained position information that allows crude tracking through the device.

The demonstrator is held in place using a support structure that is able to rotate the demonstrator components into a position aligned towards the CMS IP. The structure is constructed from a steel frame that sits on the floor of the tunnel and hosts a large bearing allowing the upper part of the structure to rotate into alignment.

The alignment of the demonstrator was performed with the help of the CERN alignment team, who installed a set of laser corner cubes that were surveyed into alignment with the CMS coordinate system. Laser cubes positioned on the detector are aligned by the CERN team with this external coordinate system. The estimated accuracy of the procedure is to point the detector to within 10 cm of the CMS IP, or about 0.1 degrees, which is more than sufficient for this experiment. The alignment is cross-checked using muons from the LHC collisions at the IP using the slabs (see Section 6).

3 Data taking

The data considered in this analysis was collected by the milliQan demonstrator between June 26th 2018 and October 21st 2018. For all physics quality data, the PMT high voltage is set to 1300 V for the R7725s, 1450 V for the R878s and 1550 V for the Electron Tubes (ETs). Events are triggered if there is a signal above the trigger threshold in three trigger groups within a window of 100 ns. Each trigger group contains two channels and the trigger group for each channel is given by floor(channel/2). The channel mapping ensures adjacent bars are in the same trigger group.

The trigger thresholds varied during data taking to satisfy rate requirements. The typical threshold for the bars ranges from $2.4 - -7$ mV, the

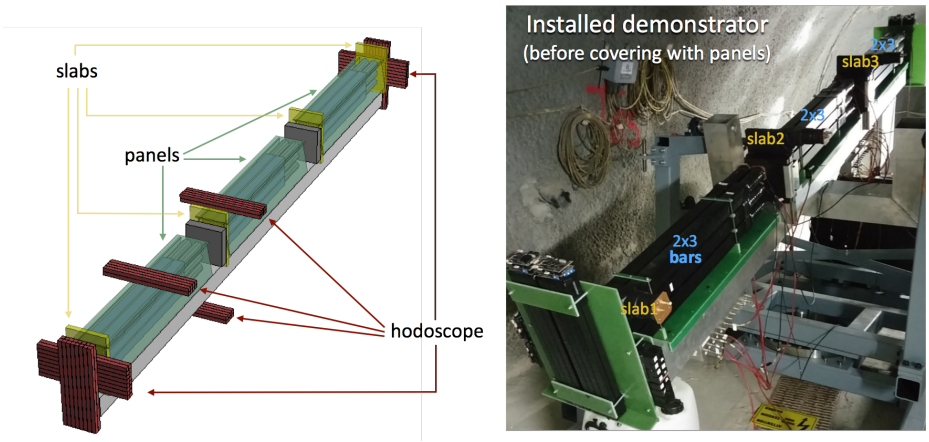


Figure 2: Design of the demonstrator and a photo of the demonstrator before covering it with panels.

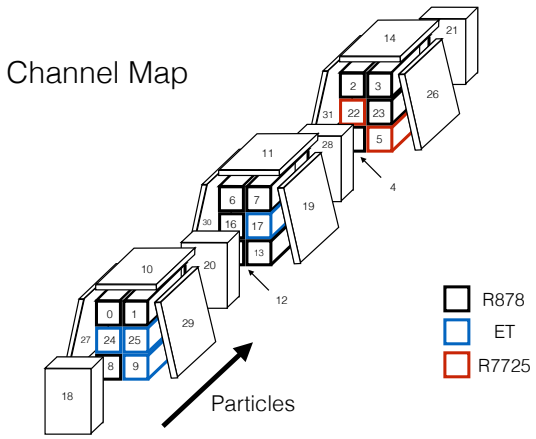


Figure 3: The channel mapping for the milliQan demonstrator. The PMT species are indicated by the colour.

typical value for the slabs ranges from $150 - 400\,\text{mV}$ and for the panels ranges from $2.5 - -40\,\text{mV}$.

The total data taking time of all physics runs is 2014.6 h. When the data taking rate is high events may not be properly synchronised between the two digitiser boards used for reading channels 0-15 and channels 16-32. To reject unsynchronised events, a requirement is made that both boards record the TDC time. The efficiency of this selection is $\sim 95\%$ giving an effective total time of 1906.4 h, with approximately 876.9 h and 1029.5 h taken during between periods with and without LHC proton-proton collisions respectively. The total proton-proton luminosity collected is $37.2\,\text{fb}^{-1}$, of which $35.1\,\text{fb}^{-1}$ satisfies the synchronisation requirement.

4 Reconstruction

The data taken by the demonstrator is comprised of "waveforms" of voltage against time for each channel. In order to reconstruct the deposits of particles in the scintillator, "pulses" are reconstructed from each waveform.

With a sampling rate of 1.6 Ghz, the waveforms span 0 to 650 ns (with the pulse that causes the trigger to accept the event placed at $\sim 360 - 390\,\text{ns}$ during triple coincidence running). The waveforms may have a significant pedestal which is measured using the mean value of the amplitude in a side-

Table 1: Mean and RMS for each channel measured in the sideband region of a physics run

Channel	Sideband mean (mV)	Sideband RMS (mV)
0	-0.2453	0.8637
1	0.009	0.7901
2	-0.007	0.9028
3	-0.07	0.7561
4	-0.661	0.8741
5	-0.252	0.7857
6	0.080	0.8706
7	-0.214	0.7795
8	-1.04	0.8098
9	-0.160	0.7954
10	1.18	0.9941
11	-0.910	0.7833
12	-0.3789	0.9148
13	-1.09	0.7899
14	-0.53	0.9600
15	–	–
16	-0.367	0.8829
17	-0.837	0.8828
18	-0.283	0.8908
19	-0.275	0.8005
20	-0.235	0.8400
21	-0.295	0.8863
22	-0.443	1.061
23	-0.606	0.8085
24	-0.762	0.9275
25	-0.201	0.8194
26	-0.576	0.9087
27	-0.440	0.7757
28	-0.706	0.8507
29	-0.426	0.9169
30	-0.280	0.9577
31	-0.458	0.8018

band region of the waveform from $0 - 50$ ns. The mean value of this pedestal measurement for each channel in a physics run is used to define a pedestal subtraction with the values shown in Table 1. Significant drifts are seen in the pedestal for channel 30 and so a subtraction is made using the sideband pedestral measurement per event.

Pulses are reconstructed from the waveforms after the pedestral subtraction. Starting from the first sample after the sideband region, the algorithm considers each time sample consecutively and runs as follows until a pulse is started:

- If the sample is above the start threshold in Table 2, increment the pulse starting counter.

- If the sample is below the reset threshold in Table 2, reset the pulse starting counter. This threshold is set as the start threshold minus half the sideband RMS shown in Table 1.
- If the pulse starting counter reaches the N samples above start threshold value shown in Table 2, start a pulse.

Once a pulse has been started the pulse end is determined by continuing as follows:

- If the sample is below the start threshold in Table 2, increment the pulse ending counter.
- If the sample is above the reset pulse end threshold in Table 2 reset the pulse ending counter. This threshold is set as the start threshold plus half the sideband RMS shown in Table 1.
- if the pulse ending counter reaches the N samples below start threshold value shown in Table 2, end the pulse.

Each pulse has several attributes defined as follows:

- time: The time of the last sample below the start threshold before the pulse is started.
- calibrated time: The calibrated time of the pulse as defined in Section 5.2.
- duration: The time of the sample after that which triggers the pulse ending minus the time of the last sample below the start threshold.
- height: The maximum amplitude in the pulse.
- area: The sum of the amplitudes across the pulse.
- N_{PE}: The number of photo-electrons (PEs) defined as the pulse area divided by the single PE (SPE) area for each channel (measured as described in Section 5.1).
- delay: Time between start of this pulse and end of previous pulse.

Table 2: Pulse finding thresholds.

N samples above start threshold	N samples below end threshold	Start threshold (mV)	Reset threshold (mV)	End threshold (mV)
7	13	2.0	1.57	2.43
7	13	2.0	1.60	2.40
9	13	2.3	1.85	2.75
7	13	2.0	1.62	2.38
7	13	2.2	1.76	2.64
5	4	2.0	1.61	2.39
7	13	2.0	1.56	2.44
7	13	2.0	1.61	2.39
7	13	2.0	1.60	2.40
4	4	2.0	1.60	2.40
7	13	2.2	1.70	2.70
7	13	2.0	1.61	2.39
7	13	2.0	1.54	2.46
8	13	2.0	1.61	2.39
9	13	2.3	1.82	2.78
7	13	2.5	2.00	3.00
7	13	2.0	1.56	2.44
4	4	2.0	1.56	2.44
8	13	2.0	1.55	2.45
7	13	2.0	1.60	2.40
7	13	2.0	1.58	2.42
8	13	2.0	1.56	2.44
5	4	3.7	3.17	4.23
7	13	2.0	1.60	2.40
4	4	2.0	1.54	2.46
4	4	2.0	1.59	2.41
7	13	2.0	1.55	2.45
7	13	2.0	1.61	2.39
7	13	2.0	1.57	2.43
7	13	2.0	1.54	2.46
7	13	2.0	1.52	2.48
7	13	2.0	1.60	2.40

5 Calibration

5.1 Charge response calibration

The calibration of the N_{PE} per unit charge incident on the detector is crucial for determining the lowest charge to which milliQan can be sensitive. To achieve this, we first compute the number of PEs for cosmic muons incident on the demonstrator. The value for N_{PE} is extracted by dividing the pulse area of cosmic muons by the pulse area of a single PE obtained from delayed scintillation PEs. The method of using delayed scintillation PEs to measure the SPE response was validated using an LED bench measurement as described in App. A. To avoid saturation effects the pulse area of

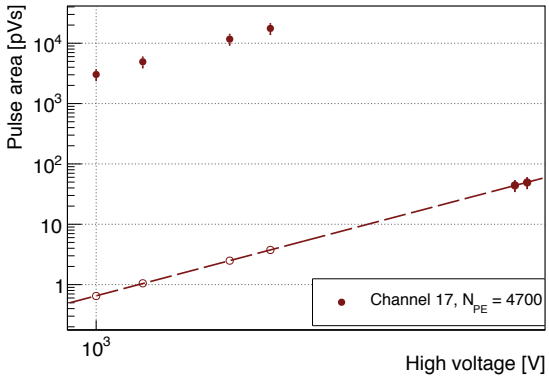


Figure 4: Area versus PMT voltage for a representative PMT for muon and single PEs. A simultaneous fit of the gain curve for the muon and SPE signals, with a floating cosmic photon yield, allows the N_{PE} for $Q = 1$ to be determined.

cosmic muons was measured at lower PMT voltage and extrapolated to the operating voltage.

Figure 4 shows the charge calibration for a representative PMT. The typical value of the measured N_{PE} for down-going cosmic muons with $Q = 1e$ is about 5000. Taking the difference in flight distance of cosmic muons and through-going muons in the scintillator bars (80 cm/5 cm), the N_{PE} for a through-going muon is approximately $5000 \times 80/5 = 80,000$. We then extrapolate N_{PE} to fractional charges by scaling by Q^2. This gives $N_{PE} = 1$ for $Q \sim 3 \times 10^{-3}e$, which is consistent with the results obtained from the full GEANT4 simulation.

5.2 Time calibration

The timing calibration is crucial to achieve the targeted $\sim$ 15 ns window between pulses arriving in each layer in order to effectively reject backgrounds from uncorrelated sources. The calibration procedure is designed such that a muon travelling upwards through the detector from the IP should have the same time value in every channel.

First, channels in the same "slice" (left or right side of each layer) are calibrated using down-going cosmic muons, tagged as a muon pulse hitting the top panel and each channel in a particular slice. The mean value of the

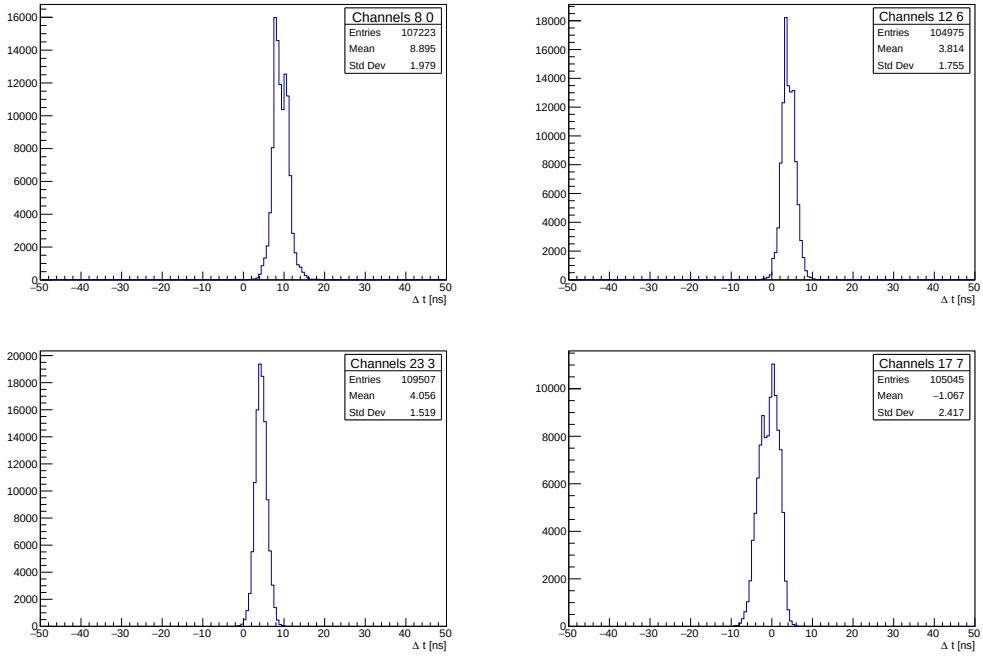


Figure 5: Example time differences between channels in the same slices. The mean values are used to calibrate the intra-slice timing.

time difference of each channel relative to the top channel of each slice is taken as the calibration. Figure 5 shows the time difference between the channels for several example calibrations. The standard deviation ranges from 1.7 to 2.8 ns.

After applying the intra-slice calibration, the slices are then calibrated using down-going cosmic muons, tagged as muon pulse hitting the top panel and both slices in a particular layer. The mean value of the time difference relative to the left slice of each layer is taken as the calibration. Figure 6 shows the time difference between the slices for each layer. The standard deviation ranges from 2.3 to 2.8 ns.

The slab timing is then calibrated using beam muons, tagged as a muon pulse hitting all four slabs. In order to avoid bias from cosmic muons, the modal value of the time difference relative to the slab closest to the CMS IP (channel 18) is taken as the calibration. Figure 7 shows the time difference between the slabs and channel 18.

After applying all previous calibrations the layers are calibrated using

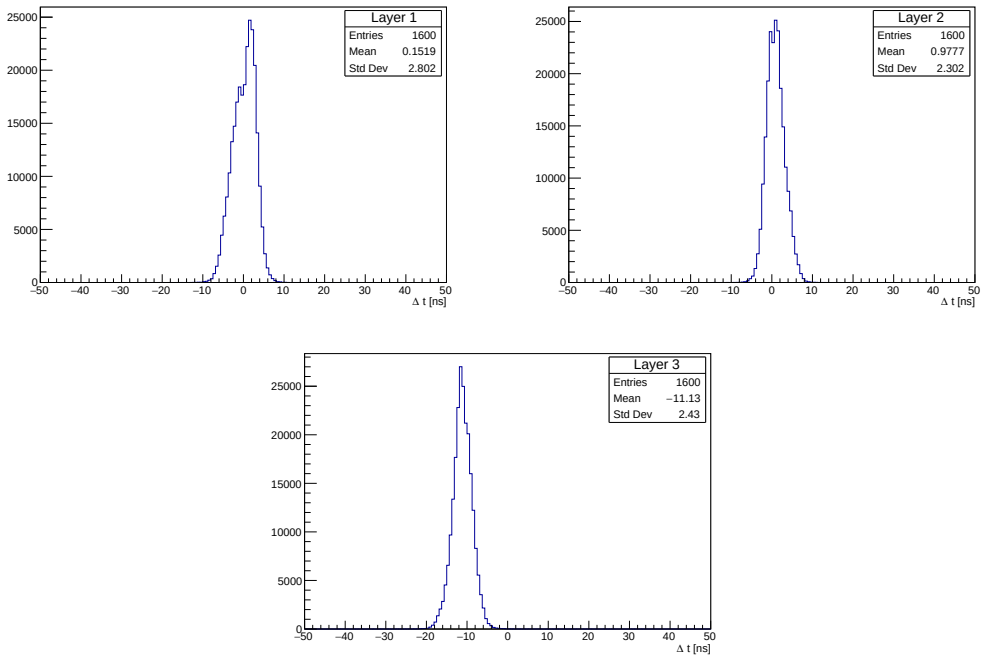


Figure 6: Time differences between slices in the same layer for all layers. The mean values are used to calibrate the intra-layer timing.

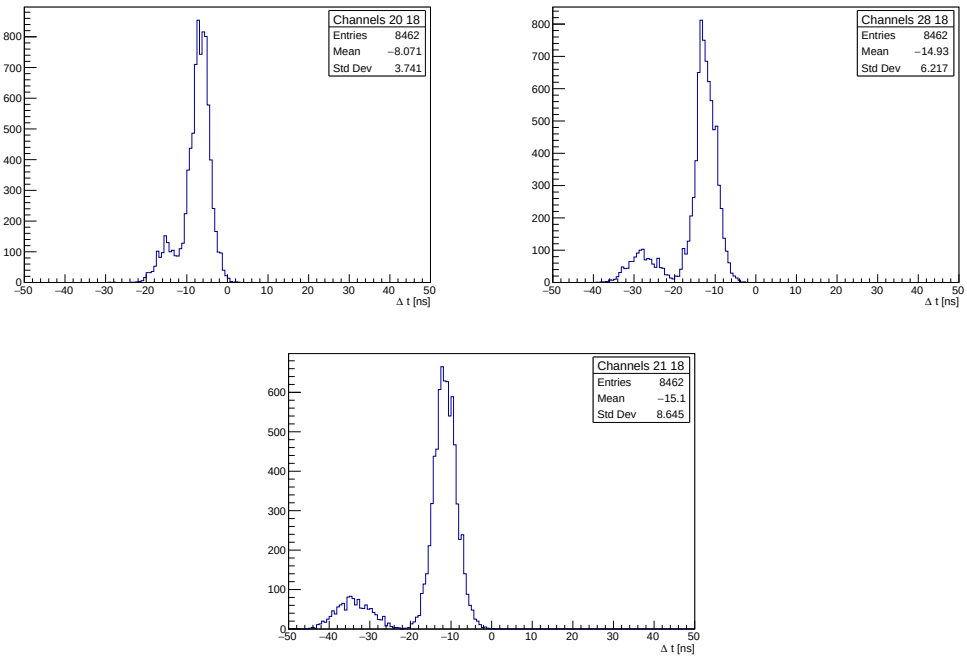


Figure 7: Time differences between each slab and channel 18. The modal values are used to calibrate the inter-slab timing.

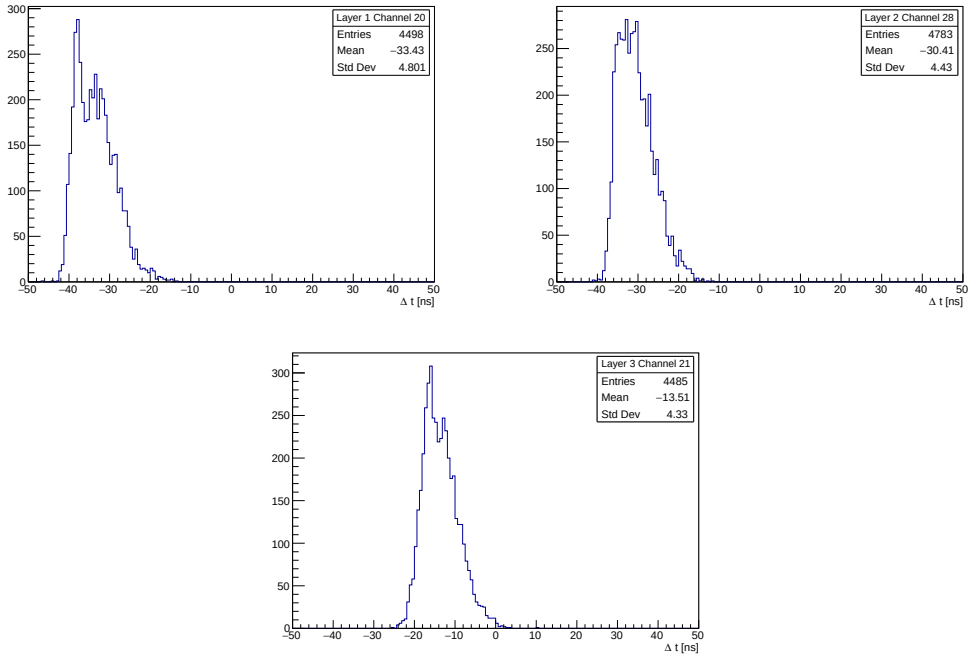


Figure 8: Example time differences between each layer and the closest slab. The mean values are used to calibrate the inter-layer timing.

beam and cosmic muons hitting all four layers, tagged as a muon pulse hitting all four slabs. The mean value of the time difference relative to the slab closest to the layer is taken as the calibration. Figure 8 shows the time difference between the layers and closest slab. The standard deviation ranges from 4.3 to 4.8 ns.

Finally, the panels are calibrated to the layers using cosmic muons tagged as having a muon pulse in the panel and in the layer closest to the panel. The mean value of the time difference relative to the layer is taken as the calibration. Figure 9 shows representative time differences between the panels and the closest layer. The standard deviation ranges from 5.8 to 8.0 ns.

The calibrations for each channel are summarised in table 3. The standard deviation in the time difference between layers is $\sim$ 5 ns, allowing a time window of 15 ns to be defined for signal candidate events, as described in Section 7.

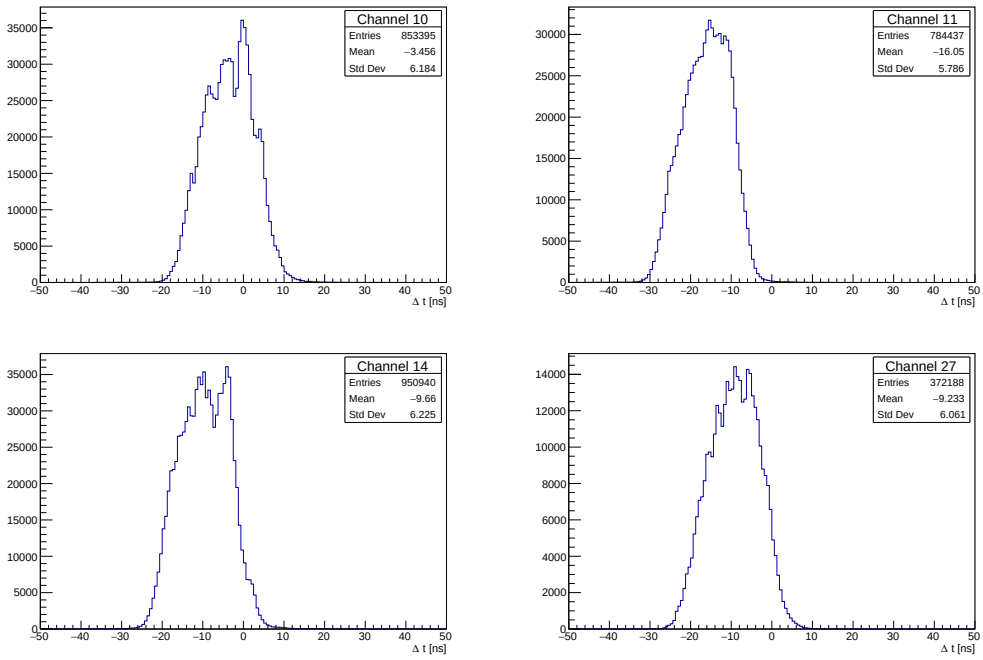


Figure 9: Example time differences between panels and the closest layer. The mean values are used to calibrate the panel timing.

Table 3: Time calibrations for each channel . The values are added to the raw time to define the calibrated time.

Channel	Calibration (ns)
0	33.125
1	33.125
2	13.75
3	24.375
4	23.75
5	35.0
6	30.625
7	29.375
8	24.375
9	33.75
10	3.75
11	16.25
12	26.875
13	34.375
14	9.375
15	–
16	27.5
17	30.625
18	0.0
19	11.25
20	7.5
21	12.5
22	28.125
23	20.625
24	33.75
25	26.875
26	-3.125
27	9.375
28	13.75
29	0.625
30	15.625
31	10.625

6 Alignment

The alignment of the demonstrator is tested using muons originating from the proton-proton collisions at CMS (referred to as "beam muons". Figure 10 shows the dependence of the total number of particles identified as having a muon pulse in all four slabs on the luminosity of the LHC fill measured by CMS. There is a clear linear dependence with a muon rate of $0.19/\,\mathrm{pb}^{-1}$ measured in data, which agrees with the expected value from simulation of a muon rate of $0.22/\,\mathrm{pb}^{-1}$. This provides confidence the demonstrator is correctly aligned with the CMS interaction point.

The time difference between the slabs closest and furthest from the IP after the calibration described in Section 5.2 also provides confirmation that

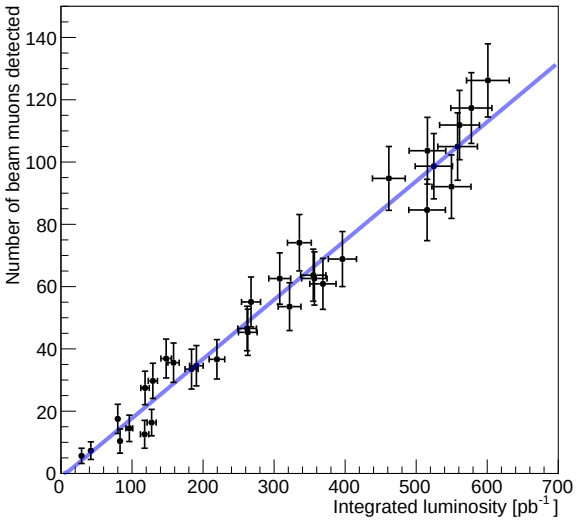


Figure 10: The occupancy of the milliQan demonstrator as a function of luminosity

the pulse timing can be used to discriminate between muons originating from the IP travelling up through the detector and muons from cosmics travelling down through the detector. Figure 11 shows well separated distributions from beam and cosmic muons with a time difference consistent with that expected from the geometrical distance between the slabs ($2 \times 3.6/0.3 = 22\,\text{ns}$).

7 Search design

In this section a search for fractionally charged particles with the milliQan demonstrator is detailed. The search relies on a triple coincidence of pulses across the three layers of the demonstrator. A range of selections are applied in order to reject contributions from background sources, which are detailed below.

- Dark rate overlap: each PMT has a dark current due to effects such as the thermal emission of electrons from the cathode. The simplest background source comes from random overlap of three such dark rate pulses. In addition, dark rate counts may overlap with a correlated

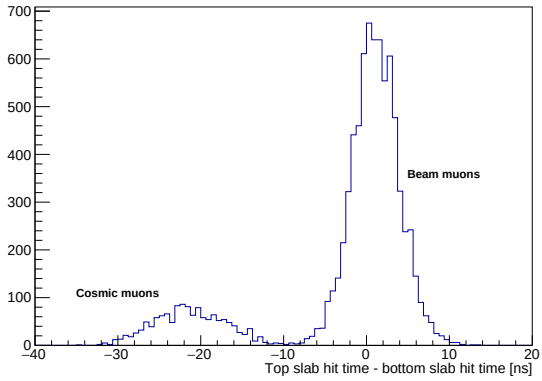


Figure 11: Time difference between the slabs furthest and closest to the CMS IP (channels 21 and 18 respectively). The peak at zero is consistent with originating from beam muons while the peak at $\sim -22\,\text{ns}$ is consistent with originating from cosmic muons.

double coincidence background from another source.

- Cosmic/beam muon showers: a large number of gammas, neutrons and electrons may be caused by an interaction of a cosmic ray muon with the rock in the demonstrator cavern. This may cause a pulse in each layer of the milliQan demonstrator. Such a background could also be expected from a beam muon which travels close to the demonstrator.

- Radiation: radiation in the cavern, scintillator bars or surrounding material can cause correlated deposits in several bars. The lead blocks placed between layers should reduce the probability of a three layer deposit arising from photons or electrons, however, neutrons will not be shielded.

- Afterpulses: afterpulses arising from correlated deposits may overlap and produce a triple coincidence signature in the demonstrator. The original correlated signature must not be triggered as in this case the afterpulses will fall in the readout deadtime and not be recorded.

Each event is required to have a pulse in a single bar in each layer in order to pass signal selection. If there is activity in any slab or panel the event is vetoed. These requirements reject backgrounds due to cosmic showers,

which are expected to cause deposits across the detector and beam muons passing close to the bars, which will cause significant pulses in the slabs. The bars which contain hits are additionally required to be pointing to the IP such that they have the same position which each layer. This reduces the background from neutrons, cosmic showers and random overlap while being efficient for signal which is expected to have a small angular spread. In each bar a requirement is made of exactly one pulse and low sideband activity to reject backgrounds from overlapping afterpulses. Finally, the maximal calibrated time difference between pulses in different bars is required to be less than 15 ns, which is efficient for signals travelling upwards from the IP and forms a powerful rejection of backgrounds with different paths through the detector or that have deposits in each layer that are uncorrelated in their timing.

8 Background determination

The total background as a function of the minimal N_{PE} in the event can be evaluated using data taking periods in which there are no collisions and scaled to the total length of the data taking with collisions. This is shown in Figure 12. The background ranges from 122 events for $0.5 < N_{PE} < 1.5$ values to 2.6 events for $N_{PE} > 10$.

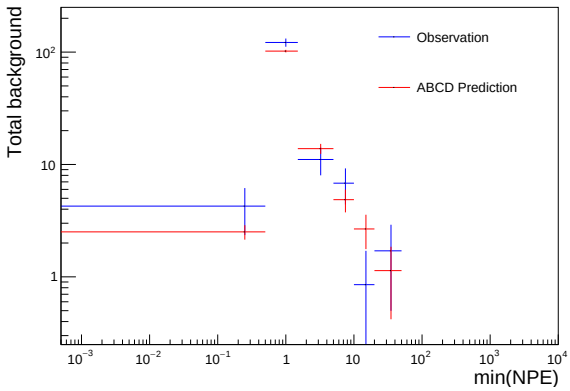


Figure 12: The minimal N_{PE} in the event during data-taking period with no collisions for events passing the signal selection is compared to the prediction from the ABCD method detailed in the text.

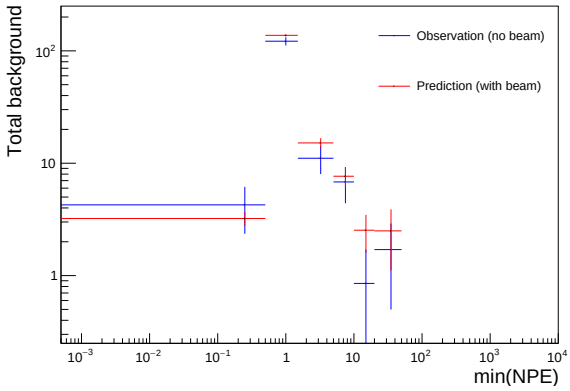


Figure 13: The minimal N_{PE} in the event during data-taking period with collisions for events passing the signal selection is compared to the prediction from the ABCD method detailed in the text.

Table 4: The region definition for the ABCD prediction.

Region	Time selection	Path selection
A (signal)	$\max(\Delta t) < 30\,\text{ns}$	Pointing path
B	$\max(\Delta t) > 30\,\text{ns}$ and $\max(\Delta t) < 100\,\text{ns}$	Pointing path
C	$\max(\Delta t) > 30\,\text{ns}$ and $\max(\Delta t) < 100\,\text{ns}$	Non-pointing path
D	$\max(\Delta t) < 30\,\text{ns}$	Non-pointing path

In order to validate there are no additional backgrounds introduced from collisions, signal depleted control regions are defined by inverting the maximal time difference and pointing path requirements. The background in the signal region is then predicted using the ABCD method for each region in N_{PE}, where the prediction follows $N_A = N_B \times N_D / N_C$ and the regions are defined in Table 4. This prediction is carried out for both the data taking periods with and without collisions and compared to the observation in the data taking period without collisions, as shown in Figures 12 and 13, respectively. Both predictions are seen to agree well with the observation, validating the use of data taken during periods without collisions to measure the background.

In order to determine the sources of the background observed in the signal region various studies have been carried out and are documented in the remainder of this section.

8.1 Dark rate overlap background

The first background source that was considered is the dark rate overlap. The dark rate is measured for each channel using "zero bias" runs in which the data is collected with a random trigger. The average dark rate for the bars is shown in Figure 14. This can be used to estimate the contribution from the random overlap of three dark counts in three bars. The prediction as a function of N_{PE} is shown in Figure 15 and compared to the observation in the signal region during data taking periods without collisions. Random overlap of dark rate can be seen to account for $\sim 25\%$ and $\sim 0.4\%$ of the overall rate for N_{PE} values of ~ 1 and > 5, respectively. The dark rates are known to vary with time and so this prediction should be repeated with a time dependent measurement.

A dark rate pulse in a particular channel may also overlap with correlated deposits in two other channels. The rate for this can be inclusively determined using runs which are collected with double coincidence triggers and then calculating the triple coincidence rate given the single coincidence rates measured as detailed above. This calculation is approximate as it does not allow for time dependence in the rates and assumes the double coincidence pulses occur at the same time. The result is shown in Fig 16. For $N_{PE} < 5$ the prediction is shown to agree reasonably well with the observation given the limitations in the method discussed above. For $N_{PE} > 5$ the prediction from this method accounts for $\sim 10\%$ of the total rate.

Finally, the background per path is shown in Figure 17. The paths involving the channels with the highest dark rates are seen to have the highest significantly higher triple coincidence rate for $N_{PE} < 5$. For $N_{PE} > 5$ the rate is comparable for all paths.

8.2 Radiation

The features of the background due to radiation in the cavern were studied using bars from the milliQan demonstrator in the cavern. First, two bars were placed side-by-side in order to mesaure the depedence of the double coincidence rate on the separation between the bars. The setup for a particular separation of 40 cm is shown in Figure 18a and the results of the rate measurement is shown in Figure **??**. For small separations (as for adjacent bars in the same layer) the rate is seen to be ~ 50 Hz. This reduces to ~ 5 Hz for a separation of 60 cm. The measurement was repeated with 12 mm of steel

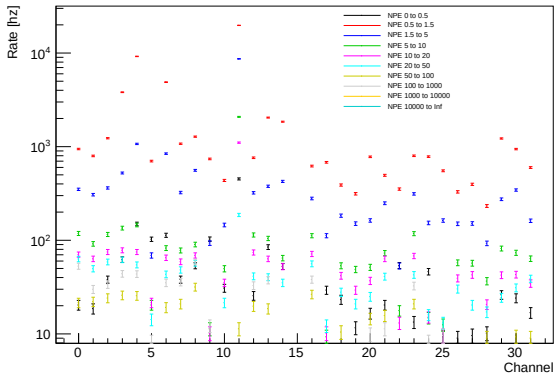


Figure 14: Single channel rate for an inclusive N_{PE} selection measured using zero bias runs.

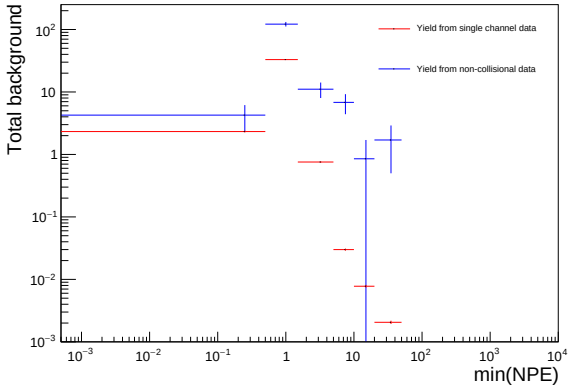


Figure 15: Background predictions for a signal like selection from dark rate is compared to the observation in the data taking period without beam.

shielding between the bars (see Fig. **??**). For all separations this is seen to reduce the overall rate by $\sim 20\%$. The time difference between the pulses in the two bars is shown in Figure 20a and Figure 20b without and with shielding, respectively, and in both cases is clearly consistent with a correlated effect. Finally, the rate against pulse area is shown in Fig. **??**. The reduction in rate is seen to be consistent across the full distribution.

The scintillator bars were placed end to end, as shown in Fig. **??**. The rate was measured for two separations (20 and 70 cm) and with lead shielding

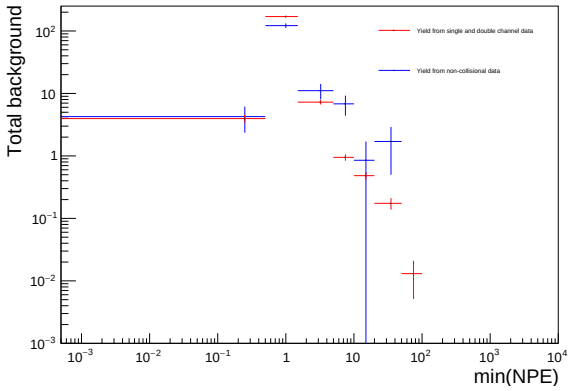


Figure 16: Background predictions for a signal like selection from the dark rate in each channel overlapping with a double coincidence in the two other channels for each path is compared to the observation in the data taking period without beam.

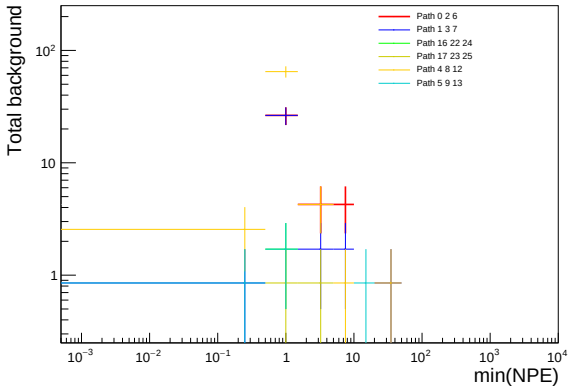
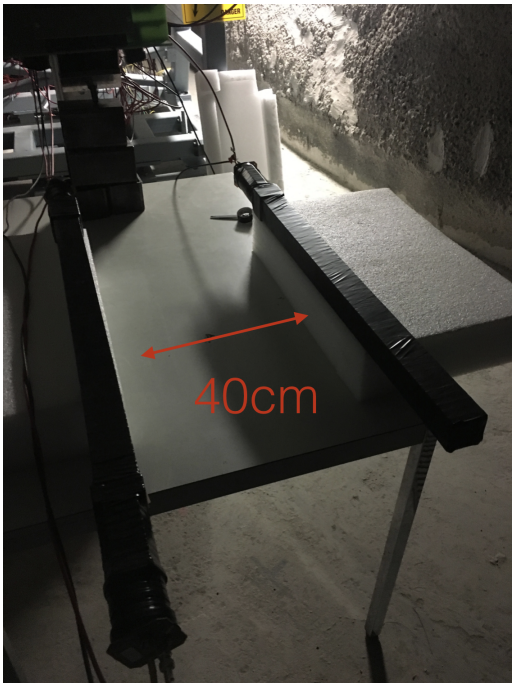


Figure 17: The minimal N_{PE} in the event during data-taking period without collisions for events passing the signal selection is shown for each pointing path through the demonstrator.

between the bars, as shown in Fig. **??**.

8.3 Cosmic/beam muon showers

To be filled ...



(a) Setup for 40 cm

(b) Setup for 2 cm with shielding

Figure 18: (a) The setup for the side-by-side rate tests with 40 cm spacing and (b) the setup for the side-by-side rate tests with 2 cm spacing and 12 mm of steel shielding.

8.4 Afterpulses

To be filled ...

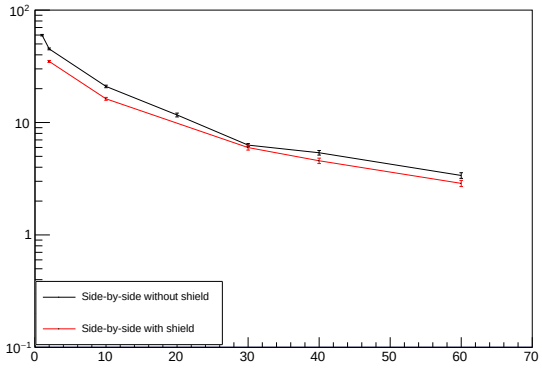
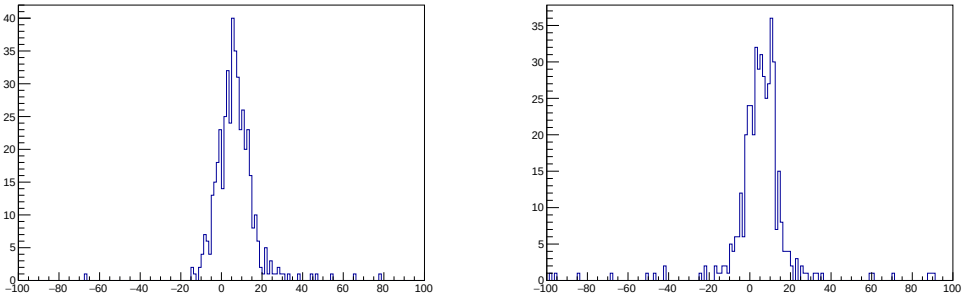


Figure 19: The rate against separation for the side by side configuration with no shielding and with shielding between the scintillator bars shown in black and red, respectively.



(a) No shielding

(b) Shielding

Figure 20: The time difference between pulses in the two scintillator bars is shown for a separation of 40 cm with shielding (a) and without shielding (b).

9 Trigger Efficiency

The trigger efficiency ε was defined as the ratio between the rate of pulses in a specific channel when linked to the trigger over the rate of pulses in the same channel when not connected to the trigger. In this way, ε represented the probability of the trigger to fire whenever a pulse was detected in the channel.

Because of the electronics available, it was not possible to compute the trigger efficiency by using a single run with different trigger configurations. ε was instead computed by comparing the trigger performance in multiple runs, each run featuring a distinct trigger configuration. These involved Triple Coincidence runs (TC) and Double Coincidence runs (DC). In the case of TC, the trigger was linked to three channels which involved two tag-channels (shared by TC and DC runs) and one probe channel, whose rate represented the numerator in the efficiency ratio. In DC runs, instead, the trigger was linked to two tag channels (these needed to be the same channels as in the TC run) and the probe channel was not connected to the trigger, hence recording all pulses indiscriminately.

According to this approach, we defined $\varepsilon = \frac{f_{TC}}{f_{DC}}$, where f_{TC} represented the pulse-rate measured in the probe channel whilst linked to the trigger (TC run), and f_{DC} represented the pulse-rate detected when the probe channel was not connected to the trigger (DC run), see figures **??**, **??**. Since different runs might span different amounts of time, instead of counting the number of pulses in each channel for each run we normalised such measure with respect to the synchronised running time. This was given by the total time of the run during which the two boards connected to channels 0-15 and 16-31 were synchronised.

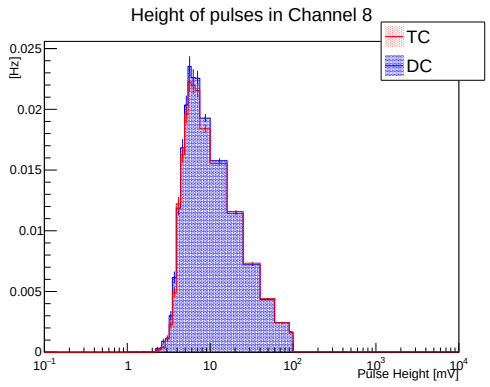


Figure 21: Comparison of pulse height rate distributions for hits in the probe channel between TC and DC runs. Channel 8 threshold at 2.4mV.

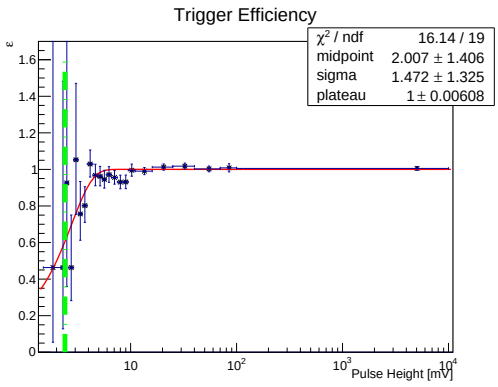


Figure 22: Trigger efficiency ε of channel 8 as a function of pulse height. ε was obtained by dividing the TC (red) bins in figure **??** by the DC (blue) bins.

9.1 Trigger Functioning

For DC runs, the trigger was armed by the first pulse in any of the two linked channels and fired whenever a second pulse was measured in the remaining channel. Similarly, for TC configurations, the trigger was armed by the first two pulses detected in any two distinct channels and was then fired when the remaining channel detected a signal. The order in which channels recorded pulses, and wether the probe channel was the one to fire the trigger or not should not affect the overall efficiency. However, care was taken to remove any potential bias in selecting events based on their timing requirements (see sections **??**, **??**).

The times associated with the online trigger were the uncalibrated times, therefore the order in which the pulses were observed by the trigger clock did not necessarily reflect their correct chronological order. When the trigger was armed, it opened a 100 ns window during which any pulse in the channels linked to the trigger could fire the trigger. If not all channels connected to the trigger measured a pulse within the 100 ns time window, the trigger would not fire and the information lost.
Whenever the first pulse crossed the trigger threshold associated with such channel, the output waveform corresponding to such pulse would be assigned a time 380 ± 10 ns, with the uncertainty introduced by the frequency with which the trigger measurements were read. Starting from the first pulse at 380ns, the remaining pulses were allocated a time based on wether they occurred before or later the arming pulse. Time differences between pulses were preserved throughout this process.

9.1.1 Timing Selection

Figures **??**, **??** show the time distribution for a comparison between TC and DC runs with tag channels 0 and 10 and probe channel 9. We observed the presence of afterpulses in the tag channels. The degree to which a channel was affected by afterpulses was connected to the trigger threshold associated with such channel. Because of the limited time window, if the threshold associated with a channel was high (usually few hundreds mVs, as in channel 0 figure **??**) the pulse activating the trigger had a longer duration, reducing the time

available in the time window for later afterpulses to occur. On the other hand, if the trigger threshold was low (few mVs, as in channel 10 figure **??**), the pulse activating the trigger was generally shorter (see, for instance, the duration of pulse in channel 24 figure **??** as opposed to the length of pulses in channels 0 or 8), leaving more time for afterpulses to occur before the 480ns cutoff. For the probe channel in figure **??**, afterpulses were removed by retaining only the earliest hit in the offline selection, but it was important to check that the current selection did not allow uncorrelated pulses in the probe channel for events that might have presented a single (uncorrelated) hit in the probe channel towards the far end of the time window.

9.2 Event Selection

We considered a time window from 280ns to 480ns (100 ns around either sides of the pulse time set by the trigger clock) within which we kept track of all the pulses detected in both the tag and probe channels. In the case of multiple hits in the tag channels, we stored the after pulses, as shown in figures **??**, **??**. If multiple hits occurred in the probe channel, we only retained the first hit above threshold (earliest pulse), disregarding the subsequent ones. We also required at least one hit in each tag and probe channel and at least one permutation of pulses (one per channel involved) to feature all pulses simultaneously within 100ns from each other.

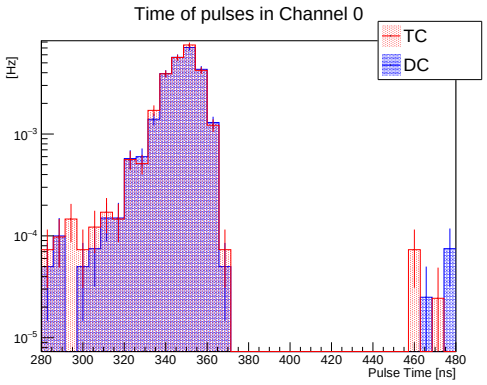


Figure 23: Pulse time distribution for hits detected in tag channel 0 at a threshold of 300mV.

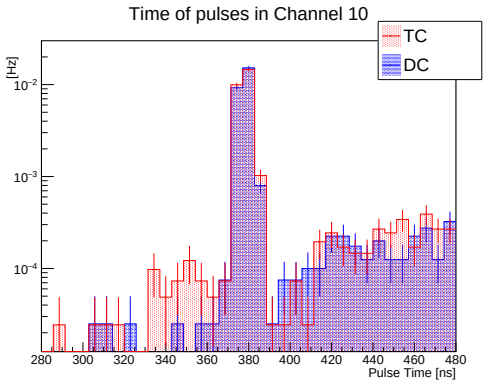


Figure 24: Pulse time distribution for hits detected in tag channel 10 at a threshold of 3mV.

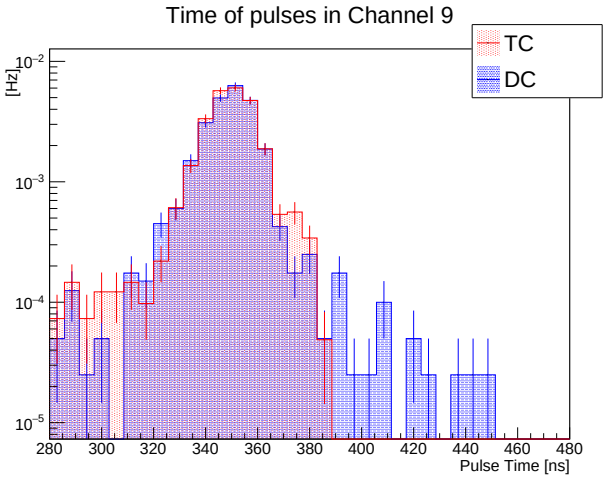


Figure 25: Pulse time distribution for hits detected in probe channel 9 at a threshold of 3.1mV.

9.2.1 Fit to the Turn On Curve

The turn on profile of the probe channel was fitted with and error function (ERF) as:

$$f_{fit} = \frac{1}{2}\lambda(1 + ERF(\frac{x-\mu}{\sqrt{2}\sigma})), \tag{1}$$

with λ plateau, μ midpoint and σ sigma (or slope) of the turn on profile, no vertical offset was included. The fit was performed according to the χ^2 method and the fit range started from 1mV below the threshold (dashed green line) until the upper limit of 10^4mV. A number of DC and SC runs was taken to test the fitting procedure by keeping the threshold of the tag channel fixed at 300mV and varying the threshold of the probe channel. This is shown in figures **??**, **??**, **??**. The threshold for the probe channel was set at 2, 3.1 and 8mV, and the midpoint of the fitting function reached larger values for higher thresholds and lower values for lower thresholds, in agreement with expectations.

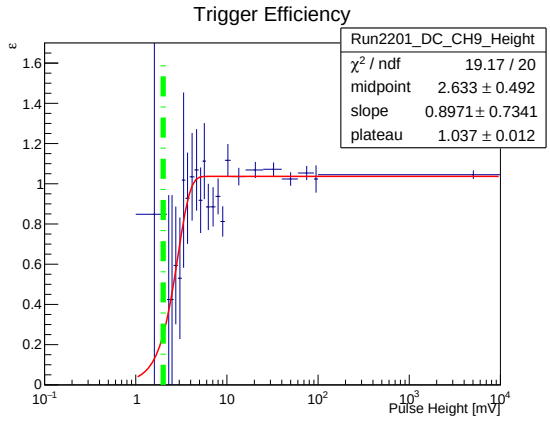


Figure 26: Turn On fit for probe channel with threshold at 2mV.

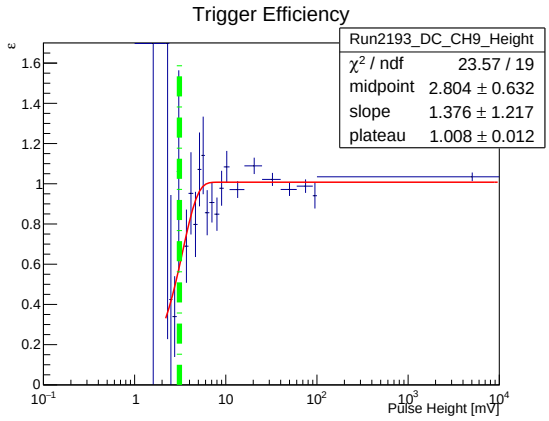


Figure 27: Turn On fit for probe channel with threshold at 3.1mV.

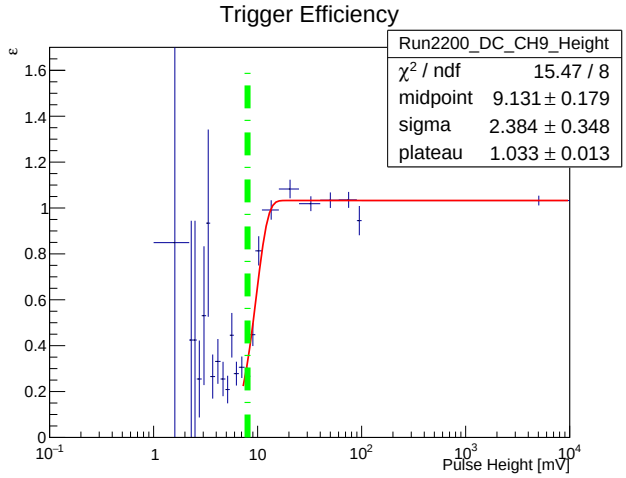


Figure 28: Turn On fit for probe channel with threshold at 8mV.

9.2.2 Rate Dependent Effects

In order to check if the rate at which the trigger was firing affected the overall trigger performance, we took runs with different tag thresholds while maintaining a constant threshold in the probe channel. Specifically, we considered DC and SC runs with tag channel 0 at 250mV (SC trigger rate 95Hz, run 2202 and 2204), 300mV (SC trigger rate 66Hz, run 2191 and 2193) and

400mV (SC trigger rate 11Hz, run 2187 and 2189), and probe channel 9 at 3.1mV. The fitted values and profiles are shown in figures **??**, **??**, **??**. Apart from the different statistics populating the histograms, no major differences were observed between run. This confirmed the expectation that the trigger efficiency would have not been sensitive to a change in frequency of the order of 100Hz or less.

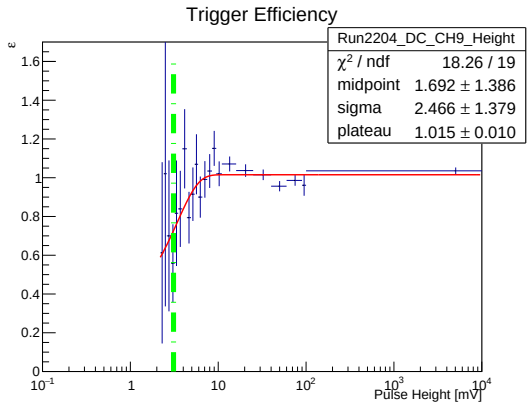


Figure 29: Trigger efficiency for probe channel 9 and tag channel 0 at 250mV.

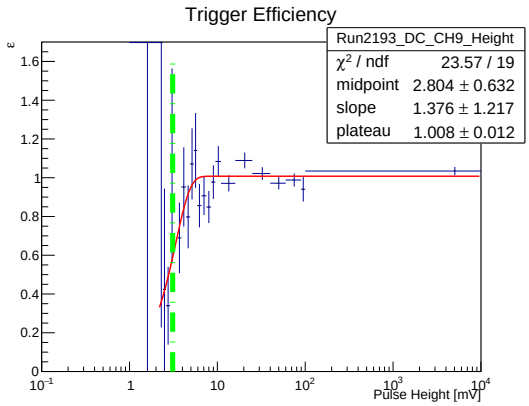


Figure 30: Trigger efficiency for probe channel 9 and tag channel 0 at 300mV.

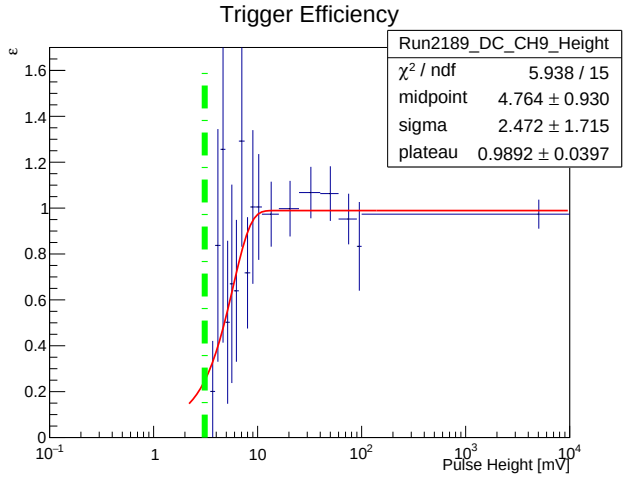


Figure 31: Trigger efficiency for probe channel 9 and tag channel 0 at 400mV.

9.2.3 Single Coincidence Runs

To gain a better understanding of how the trigger worked, we also made use of Single Coincidence runs (SC) in which the trigger fired on any pulse of a designated tag channel. In order to maintain the trigger rate low (below 100Hz), the threshold of the associated SC channel was set to larger values (typically 300-400mV). This introduced a delay in the output waveform times assigned by the trigger clock, since the point in time where the pulse crosses the thresholds is set around 380ns, the start of the pulse itself typically occurs 25-30ns before (for trigger threshold of about 300mV). The ROOT-tree data used in this analysis defined the pulse time as the beginning of the pulse, instead of the time time when the signal crossed the threshold. This feature is outlined in figures **??**, **??**.

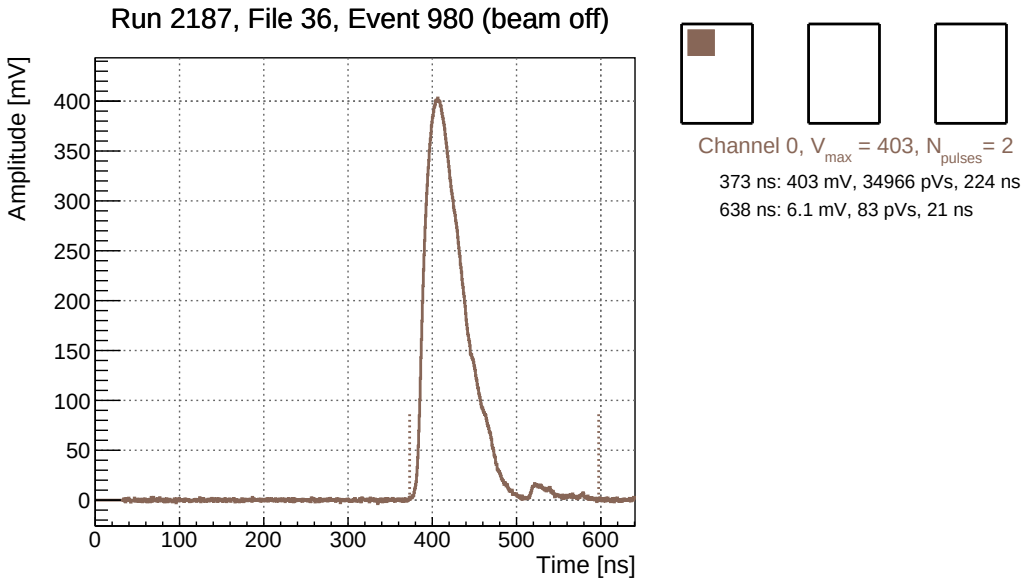


Figure 32: Event display of a SC run with trigger threshold set on channel 0 at 400mV. As shown in the diagram, the time difference between the start of the pulse (around 373ns) and the point where it crosses threshold (around 400ns) determines by how much the waveform will be time-translated in the offline data.

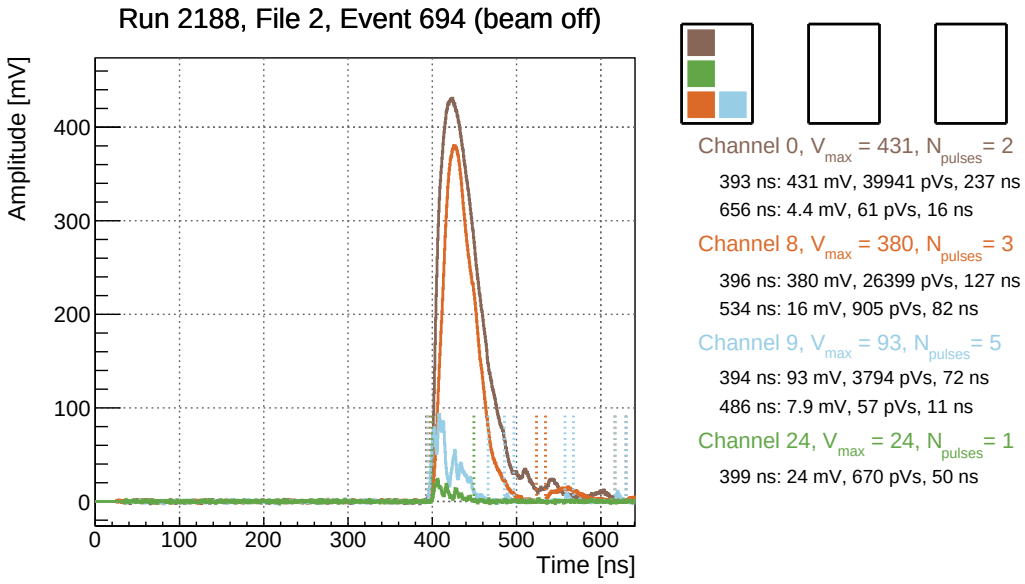


Figure 33: Event display of a DC run with trigger thresholds set on channel 0 at 400mV and on channel 8 at 2.4mV. In this current trigger configuration the pulse in channel 8 almost immediately crosses the (low) trigger threshold, and the trigger will set both pulses only few ns apart from the 380ns time-mark, as opposed to the $\sim$ 30ns delay introduced by the pulse in fig **??**.

9.3 Efficiency for Individual Channels

The following plots represent the trigger efficiency for some of the channels in the first layer as a function of pulse height in mV. Only 4 channels were considered since longer runs (20 hours) were needed in order to increase statistics.

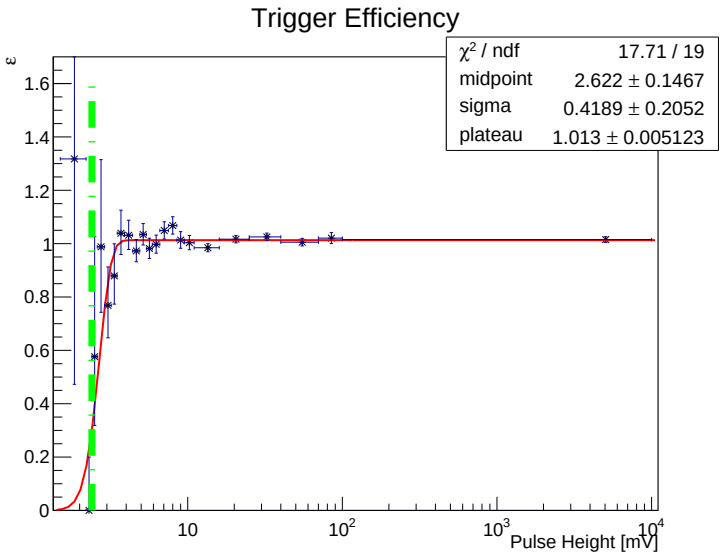


Figure 34: Trigger efficiency for channel 0 as a function of pulse height.

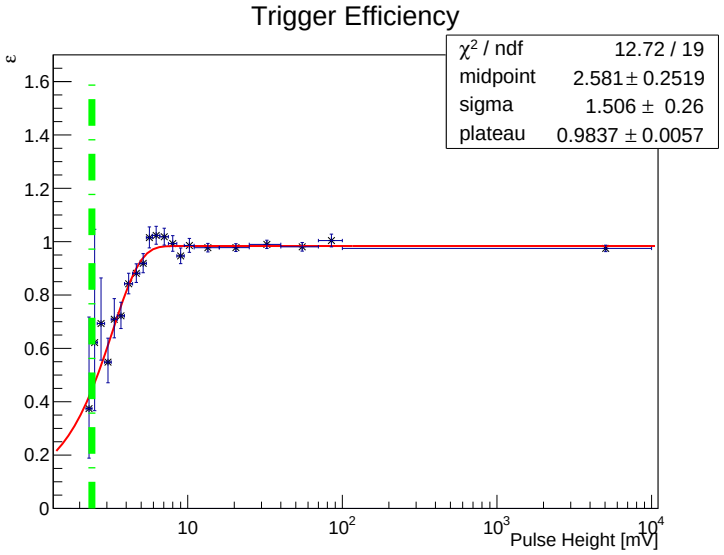


Figure 35: Trigger efficiency for channel 1 as a function of pulse height.

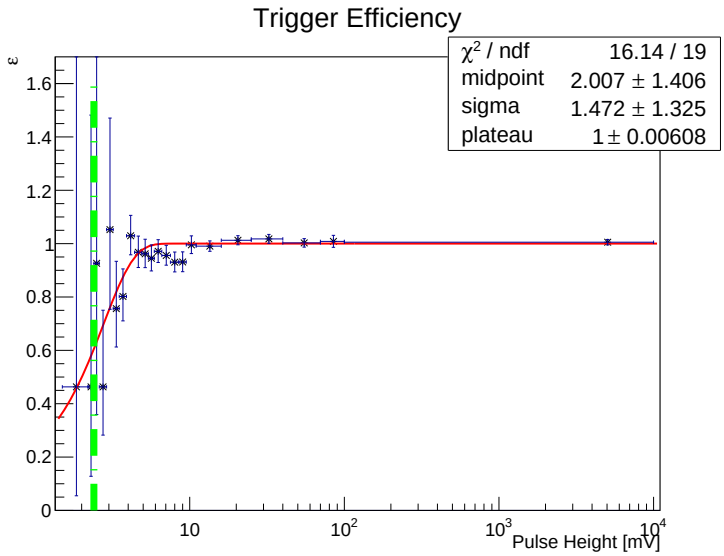


Figure 36: Trigger efficiency for channel 8 as a function of pulse height.

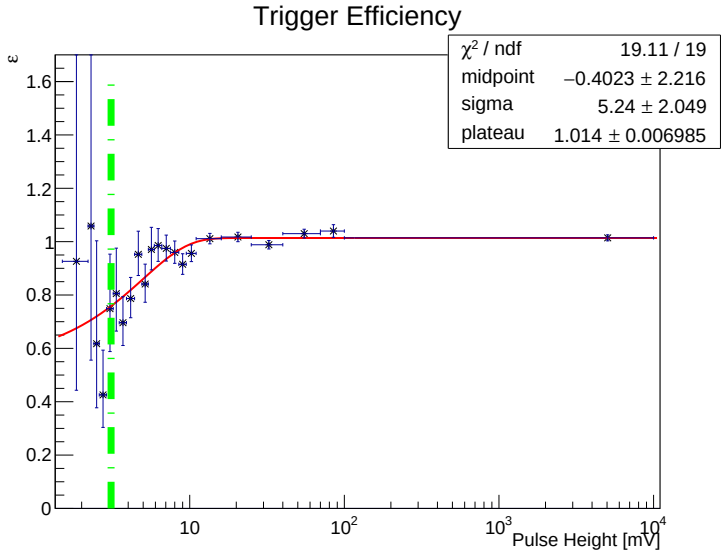


Figure 37: Trigger efficiency for channel 9 as a function of pulse height.

10 Interpretation

An initial approximation of the exclusion limits at 95% CL is shown in Fig. 21. This uses the legacy simulation and signal generation and propagation. Three

"signal regions" are then defined for charges of $Q < 0.01$, $0.01 < Q < 0.02$ and $Q > 0.02$ with background yields, measured in Figure 12, taken from the regions $N_{PE} < 5$, $5 < N_{PE} < 20$ and $N_{PE} > 20$, respectively. The reach is shown to extend to cover masses between ~ 0.1 and 5 GeV and charges between ~ 0.02 and $0.1e$.

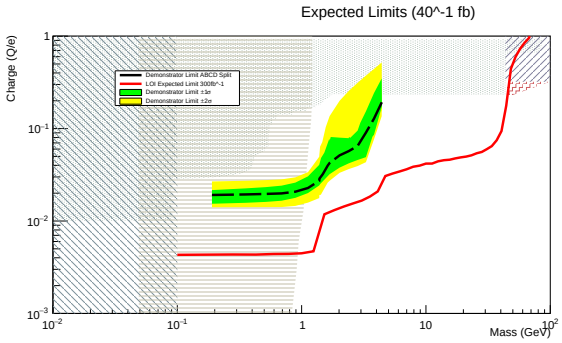


Figure 38: The contours of 95% CL upper limits are shown for the demonstrator and full milliQan detector.

11 Summary

The alignment and calibration of the milliQan demonstrator have been detailed. The calibrated data have been used to search for fractionally charged particles with new exclusion sensitivity expected for masses between ~ 0.1 and 5 GeV and charges between ~ 0.02 and $0.1e$.

A LED bench studies

In this appendix the studies undertaken using PMTs coupled to an LED source are documented. The experimental setup is shown in Figure 22. The LED is a Thorlabs LED430L with a 430 nm wavelength and representative R878, R7725 and ET PMTs are studied. The DRS scope is triggered on the LED pulse meaning the PMT response falls in a well-defined time window and removing the need for pulse finding. The pulse areas for a range of R878 HVs are shown in Figure 23. The average N_{PE} input to the PMT can be controlled by varying the amplitude of the input LED pulse. This setup was

used to study the SPE area, afterpulses and create an SPE template for pulse and signal injection tests.

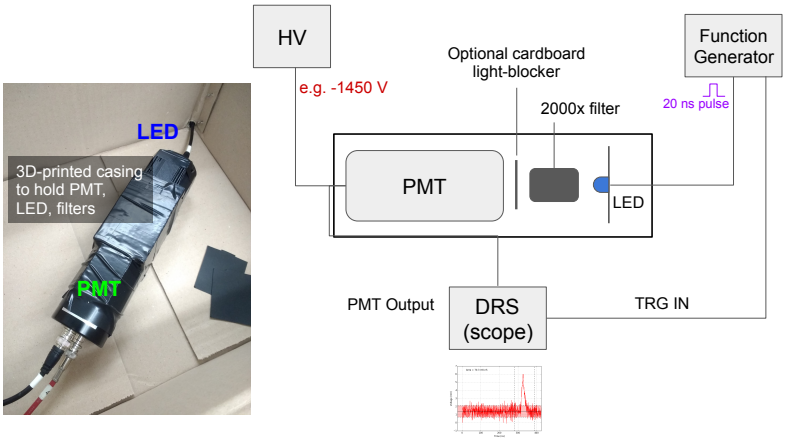


Figure 39: The experimental setup used for the LED bench studies.

A.1 SPE area measurement

The calibration of the SPE area using an LED allows the impact of "partial" SPE pulses to be considered. Such pulses can occur in the PMT due to non-optimal trajectories such as the photon passing through the cathode and directly striking the first dynode or a PE skipping a dynode. The calibration procedure follows the method described in Ref. [1]. First, the zero SPE pulse area distribution is measured by taking data with the LED blocked. Next, the pulse area distribution is measured with the LED on and the zero SPE distribution scaled to match the left edge of this distribution. This is shown in Figure 24. The fraction of zero SPE events to be estimated and, assuming a Poisson distribution, the average N_{PE} from the LED can be calculated as $\langle N_{PE} \rangle = -\log(f)$.

The average SPE area can then be trivially calculated as $\langle \text{area}_{\text{on}} \rangle - \langle \text{area}_{\text{blocked}} \rangle / \langle N_{PE} \rangle$. The result of this procedure is shown in Figure 25 for

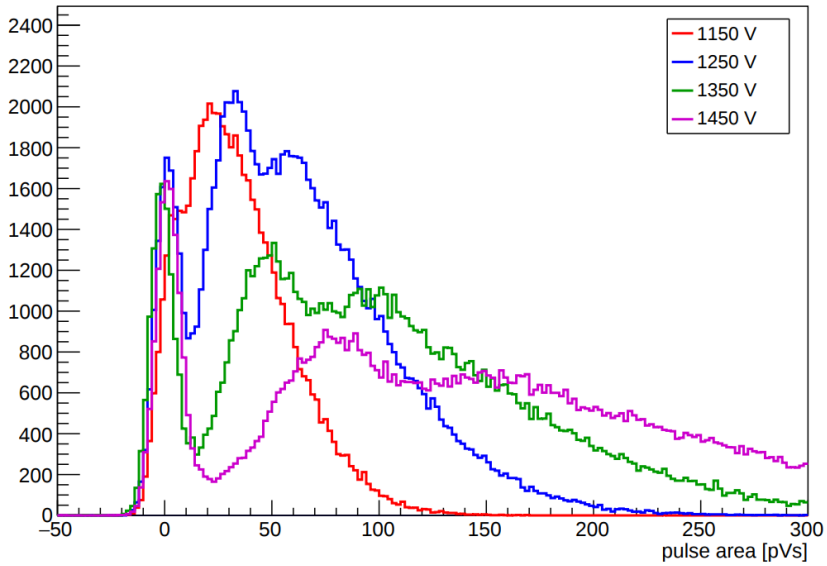


Figure 40: The area distributions for several R878 PMT HVs. The peaks correspond to 0, 1 and 2 SPEs and their separation increases with HV.

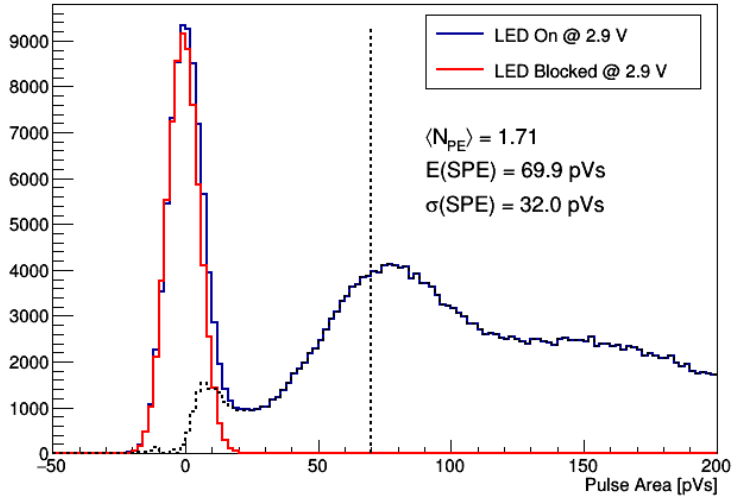
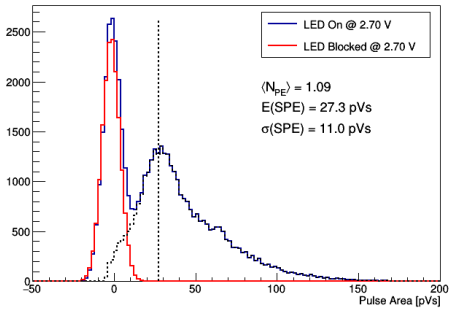


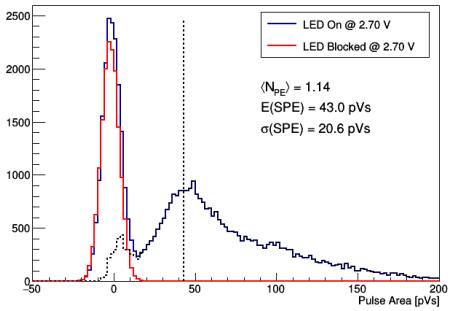
Figure 41: The pulse area distributions used to calibrate the SPE area for HV = 1450 V.

a range of HV values. The measured value for 1450 V is in good agreement with the in-situ measurement of the R878s described in Section 5.1.

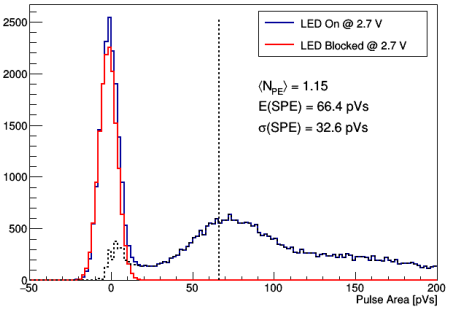
The measurement of the SPE area for an R7725 and an ET PMT is shown in Figure 26 and 27 respectively. In the case of the R7725 the SPE area is significantly larger compared to the R878 and has a larger proportion of partial SPEs causing the mean response to be $\sim$ 20% below the peak. In addition there is a significant disagreement with the in-situ measurement described in Section 5.1. This is likely due to the in-situ method neglecting the partial SPEs and the magnetic field in the cavern reducing the response by causing photoelectrons to be diverted.



(a) 1250 V

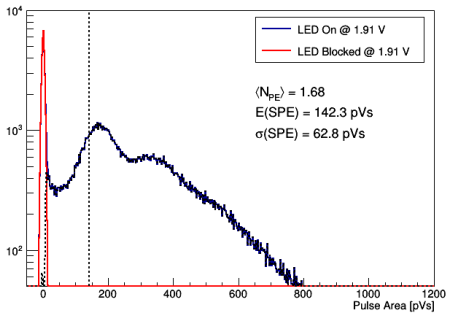


(b) 1350 V

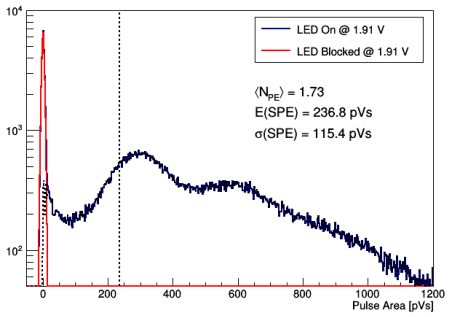


(c) 1450 V

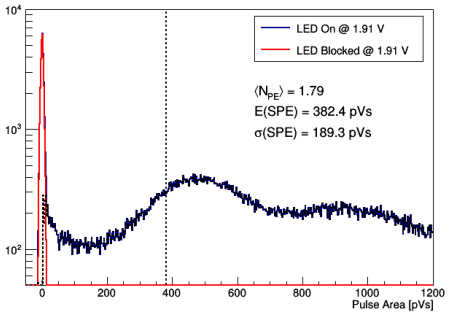
Figure 42: The pulse area distributions used to calibrate the SPE area of an R878 PMT for a range of HV values.



(a) 1250 V



(b) 1350 V



(c) 1450 V

Figure 43: The pulse area distributions used to calibrate the SPE area of an R7725 PMT for a range of HV values.

A.2 Afterpulse characterisation

Afterpulses are caused by the production of positive ions in residual gases in the PMT tube. These ions return to the photocathode and produce a significant number of photoelectrons. The amplitude and time delay depend on the mass and charge of the ion as well as the position of the production of the ion. The time delay typically ranges from several hundred nanoseconds to a few microseconds or more. In the later case this can cause a correlated background in the detector if there are overlapping afterpulses from a single correlated source (which would lie outside of the trigger window).

The afterpulse behaviour was studied for the R878 and R7725 PMT considered in Section A.1. The time window was expanded to a microsecond

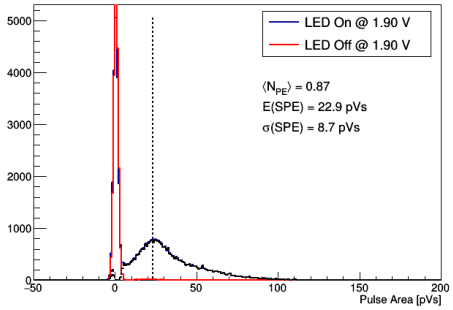
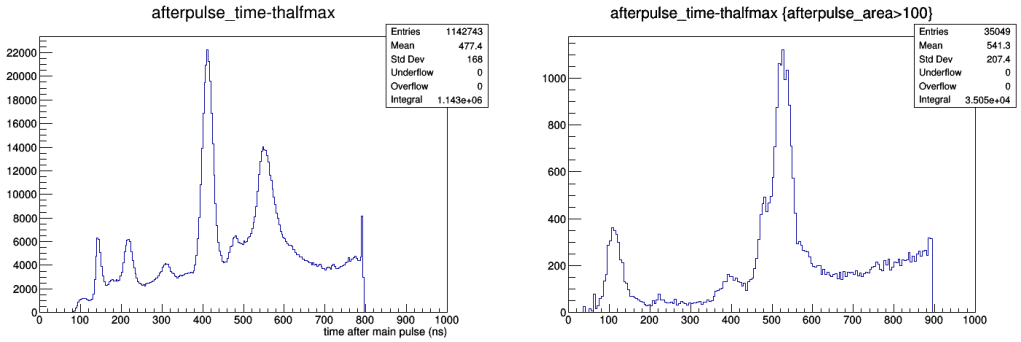


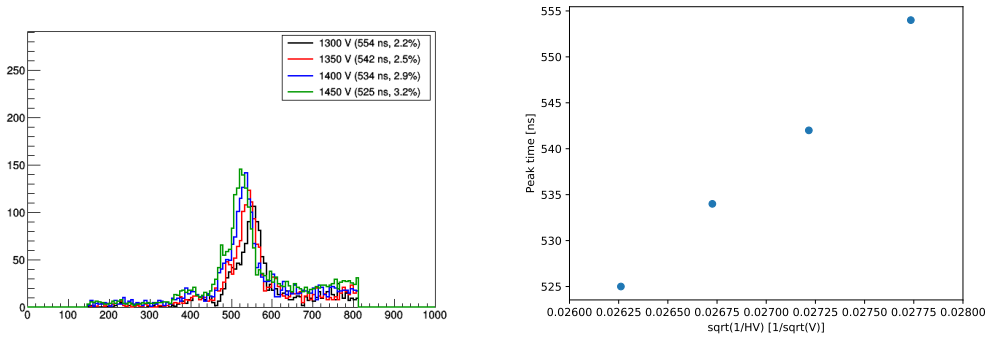
Figure 44: The pulse area distributions used to calibrate the SPE area of an ET PMT for a HV of 1700 V.

and the afterpulse spectrum measured by averaging the waveforms after subtracting the time of the initial pulse. The results are shown in Figure 28 for the R878 and R7725 PMTs. The R878 shows several peaks with the largest occuring at $\sim$ 400 ns while the R7725 has a single large peak at $\sim$ 500 ns. Finally, the average waveform for various HV values is shown in Figure 29a. The peak time is expected to follow an inverse square root HV dependance [2] and, while this requires further study, can be seen in Figure 29b.



(a) R878 (b) R7725

Figure 45: The average waveform spectrum for two PMTs. The peaks correspond to afterpulses.

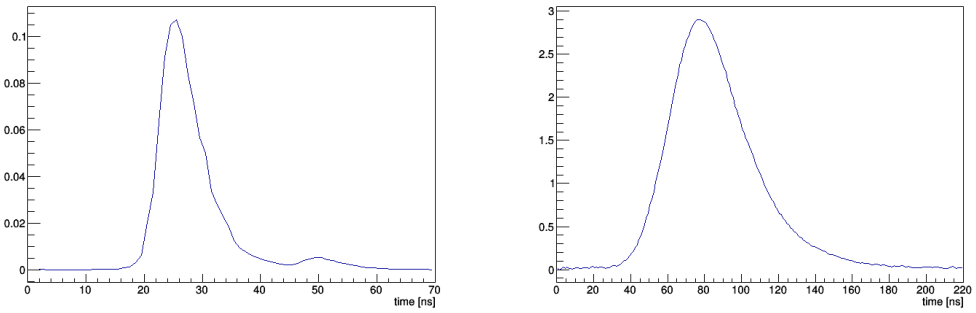


(a) (b)

Figure 46: (a) The average waveform spectrum for the R7725 PMT for several HV values. (b) The peak time follows an inverse square root dependence with the HV.

A.3 SPE pulse shape and area spectrum

The LED is finally used to measure the average SPE waveform and are spectrum. This is carried out for an R878 and R7725 PMT. The pulse shape is measured by averaging many individual pulses and the resultant templates are shown in Figure 30. The average waveform for the R7725 is significantly narrower than for the R878. The SPE area spectrum is measured by using a very low LED intensity to eliminate contamination from 2 or more SPE events. The zero PE component is then subtracted following the same procedure in A.1 and the resultant SPE area distribution for the R7725 shown in Figure 31. There is an exponential tail from suboptimal trajectories at low pulse areas and a Gaussian core around $\sim 500\,\mathrm{ps}$. The contamination from 2 or more SPE events is $\sim 3\%$.



(a) R878 (b) R7725

Figure 47: The average waveform spectrum for the R7725 PMT for several HV values. The peak time follows an inverse square root de dependence pendance with the HV.

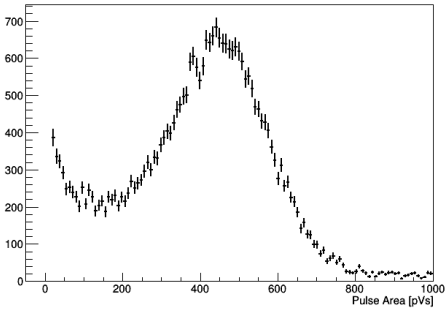


Figure 48: The measurement of the SPE area spectrum for an R7725 PMT.

References

[1] R. Saldanha, L. Grandi, Y. Guardincerri, and T. Wester. Model Independent Approach to the Single Photoelectron Calibration of Photomultiplier Tubes. *Nucl. Instrum. Meth.*, A863:35, 2017.

[2] K.J. Ma, W.G. Kang, J.K. Ahn, S. Choi, Y. Choi, M.J. Hwang, J.S. Jang, E.J. Jeon, K.K. Joo, H.S. Kim, J.Y. Kim, S.B. Kim, S.H. Kim, W. Kim, Y.D. Kim, J. Lee, I.T. Lim, Y.D. Oh, M.Y. Pac, C.W. Park, I.G. Park, K.S. Park, S.S. Stepanyan, and I. Yu. Time and amplitude of afterpulse measured with a large size photomultiplier tube. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 629(1):93, 2011.